PAPER • **OPEN ACCESS**

Entropy-probabilistic modeling as a tool for forming key competencies of a doctor

To cite this article: A N Tyrsin *et al* 2020 *J. Phys.: Conf. Ser.* **1691** 012162

Entropy-probabilistic modeling as a tool for forming key competencies of a doctor

A N Tyrsin[1,3], D A Yashin[2] and A A Surina[3]

[1] Ural Federal University, 19, Mira str., Yekaterinburg, 620002, Russian Federation
[2] South Ural State Medical University, 64, Vorovskogo str., Chelyabinsk, 454092, Russian Federation
[3] South Ural State University, 76, Lenin ave., Chelyabinsk, 454080, Russian Federation

E-mail: at2001@yandex.ru

Abstract. The possibility of using modern digital educational technologies in the formation of key competencies of a doctor is shown on the example of the entropy-probability model, which is a synthesis of the system-entropy approach and multidimensional risk analysis of stochastic systems. Examples of practical application of this model in preventive medicine are given: analysis of population entropy in the prevention of noncommunicable diseases, comprehensive assessment of the effectiveness and safety of medicines, quantitative assessment of population health with the determination of the contribution of individual risk factors, and systematic analysis of population changes in monitoring risk factors. The introduction of entropy-probabilistic modeling in the educational process will help in the formation of the doctor's basic universal, professional competencies and systematic clinical thinking.

1. Introduction

In the professional activity of a doctor, the main objects of research and influence are people and the population which can be interpreted as large systems of many interacting parts. All this variety of interdependent subsystems must be taken into account by the doctor in his daily clinical work – you can not stop only on the treatment of a particular disease. The most important principle of clinical medicine – "Treat the patient, not the disease", put forward at the beginning of the XIX century by the famous Russian Clinician M. Ya. Mudrov [1], is formed on the basis and using a systematic approach, and remains relevant at the present stage, which is reflected, among other things, in the development of personalized medicine [2].

When preventing diseases, it is important to consider the impact of several risk factors (FR), both directly and in interaction with each other. It is important to correctly assess the state of health based on objective quantitative indicators. The need to develop a doctor's knowledge, skills and system analysis skills is reflected in the Federal state educational standards of higher education (FSES HE) for the relevant medical specialties. For example, in accordance with the FSES HE in the specialty "General medical practice (family medicine)", a graduate should have: universal competence No. 1 – "readiness for abstract thinking, analysis, synthesis" and other key professional competencies (PC) formed using a systematic approach (for example, PC 1, 2, 4, etc.) [3]. However, in educational programs in the specialty "Medical science" (specialty level) – this is the specialty that most graduates graduate from medical universities residency programs, methods and technologies of multidimensional system analysis and management of stochastic complex systems are not sufficiently represented. We believe that entropy-probabilistic modeling is a promising tool for the formation of key competencies of a doctor.

2. The entropy-probability model

The entropy-probability model is based on system-entropy approach and multidimensional risk analysis of stochastic systems.

System analysis is based on the representation of a population, a person (as well as its components – a cell, an organ, a functional system, etc.) as a complex, multidimensional, probabilistic (stochastic), open and self-organizing system. A system is "an integral formation consisting of a set of components, the interaction of which causes the appearance of new, integrative qualities that are not peculiar to the parts that form it" [4]. Another definition of a biological system is a set of interrelated and dependent elements that, forming a single whole, perform certain functions, as well as interact with the environment or other elements and systems [5].

The main characteristics of the system include: the appearance of new properties in it in comparison with the elements that it consists of; the existence of the system as an organizational unity of elements (violation of relationships leads to the destruction of the system); systems exist and are created to solve certain problems. The provision on the appearance of new integrative qualities that are not characteristic of individual components of the system [6] is important in preventive medicine. An example of this new quality is comorbidity [7], immunization of the population in the prevention of infectious diseases, etc. In medicine, the system is (on the example of a population as a preventive medical and biosocial system): multidimensional, whose components can be the main biological FR (dyslipidemia, hypertension, overweight, hyperglycemia), behavioral FR (Smoking, unhealthy diet, inactivity, harmful alcohol consumption) [8], major non-communicable diseases (NCDS) , etc.; stochastic, i.e. with probabilistic behavior; open (exchanging energy and information with the environment); self-organizing (able to adapt to changing environmental conditions); complex.

Entropy [9] is reasonably used to study systems of the class under consideration [10]. An object is described as a system defined by a vector $\mathbf{Y} = (Y_1, Y_2, .., Y_m)$, in which each component Y_i determines the evaluation of a specific subsystem of the object. The entropy calculation is based on the formula [11]

$$\mathrm{H}(\mathbf{Y}) = \mathrm{H}(\mathbf{Y})_\mathrm{V} + \mathrm{H}(\mathbf{Y})_\mathrm{R} = \sum_{i=1}^{m} \ln \sigma_{Y_i} + \sum_{i=1}^{m} \kappa_i + \frac{1}{2}\sum_{k=2}^{m} \ln(1 - R^2_{Y_k/Y_1 Y_2 \ldots Y_{k-1}}),$$

where $H(\mathbf{Y})$ – general entropy; $H(\mathbf{Y})v$ – entropy of randomness.

$H(\mathbf{Y})_\sigma = \sum_{i=1}^{m} \ln \sigma_{Y_i}$ – randomness entropy of scattering. $H(\mathbf{Y})_\kappa = \sum_{i=1}^{m} \kappa_i$ – randomness entropy. $\kappa_i = H(Y_i / \sigma_{Y_i})$ – entropy indicator of the random variable Y_i type of distribution law. $H(\mathbf{Y})_R = \frac{1}{2}\sum_{k=2}^{m} \ln(1 - R^2_{Y_k/Y_1 Y_2 \ldots Y_{k-1}})$ – entropy of self-organization - reflects the mutual dependence of all elements of the system $\mathbf{Y}$.

3. Methodology of the system-entropy analysis technique

The method of system entropy analysis (SEA) was used in disease prevention to analyze the entropy of the population (n=1402 people) in groups of healthy, practically healthy and sick [10]. As the health status of the population aged 25-34 years worsens, the total entropy increases: in the group of healthy people – 5.52, in practically healthy people – 7.37, and in patients – 8.08. The practical significance of this fact is as follows: if several surveyed populations have a higher level of entropy in one of them, then without further examination, we can first say about the worst state of health of this group and allocate it for a more in-depth examination and priority prevention. Also, based on the analysis of determination indices, it is possible to determine the contribution of each subsystem to the change in entropy.

SEA allows us to provide a comprehensive numerical assessment of population health at a qualitatively new level using different types of data, both quantitative and qualitative. SEA becomes particularly important in cases where a preliminary assessment of the results of a study with a limited sample is required, or when even with a sufficient number of observations, a small range of differences does not allow making an unambiguous decision in favor of a particular drug (method, etc.) when using generally accepted methods of statistical analysis. This possibility is particularly relevant at the moment, when in the context of the new COVID-19 coronavirus pandemic, millions of patients receive off-label treatment based on preliminary data until all stages of clinical trials are completed in accordance with the criteria of evidence-based medicine [12, 13].

SEA has been used in clinical medicine for a comprehensive assessment of the effectiveness and safety of medicines on the example of the use of lipid-normalizing drugs (LDL) in the treatment of patients with CHD [14]. The sample included 219 people from a clinical group of patients who participated in an open study on the effectiveness and safety of statins (lovastatin, fluvastatin, simvastatin, atorvastatin), fibrates (ciprofibrate). This study Evaluated the effect of LDL on the main components of the lipid system that are involved in atherogenesis: the subsystems “Total cholesterol”, ”Triglycerides”, ”HDL Cholesterol”, i.e., the most accessible laboratory-determined components of the lipidogram.

According to the “Hypo-entropic” effect, LDL was distributed as follows (in descending order of the dynamics of the total entropy $\Delta H(\mathbf{Y})$): atorvastatin > fluvastatin > ciprofibrate > simvastatin > lovastatin. When treated with atorvastatin $\Delta H(\mathbf{Y})=\mathbf{-1{,}495}$ (decrease from 3.3 to 1.8 by 45.3%). The decrease in total entropy was mainly due to a decrease in the entropy of randomness. According to the results of a comprehensive analysis using sea, atorvastatin is the most effective lipid-normalizing drug, which corresponds to the results of other studies [15].

SEA can be used to assess the safety of medicines, especially for small samples, using the calculation of the entropy multidimensional cross-correlation relationship between all elements of subsystems before and after treatment (or between different systems), i.e. through the calculation of the entropy of the relationship [16]. Let **X** and **Y** be continuous random vectors. Then the entropy of the relationship between **X** and **Y** is equal to

$$H(\mathbf{X}\cap\mathbf{Y}) = -\frac{1}{2}\ln\left(1-\frac{1-d_e(\mathbf{Z})}{(1-d_e(\mathbf{X}))(1-d_e(\mathbf{Y}))}\right),$$

where $d_e(\mathbf{X})$, $d_e(\mathbf{Y})$, $d_e(\mathbf{Z})$ – the coefficients of closeness of the joint correlation **X**, **Y**, **Z** [9].

If $\mathbf{X}^\circ$ and $\mathbf{Y}^\circ$ – Gaussian random vectors, when $H(\mathbf{X}^\circ\cap\mathbf{Y}^\circ) = -\frac{1}{2}\ln\frac{\left|\mathbf{R}_{\mathbf{X}^\circ\cup\mathbf{Y}^\circ}\right|}{\left|\mathbf{R}_{\mathbf{X}^\bullet}\right|\cdot\left|\mathbf{R}_{\mathbf{Y}^\bullet}\right|}$.

4. An example of the analysis of drug safety

As an example of drug safety analysis, we used an assessment of the multidimensional entropy system relationship between elements of the lipid system (**X**) and liver function indicators (total bilirubin, AST and ALT enzymes) (**Y**), which change when a hepatotoxic effect is manifested (table 1) during statin therapy. The result can also be evaluated between one system before and after treatment. $H(\mathbf{X}_0\cap\mathbf{Y}_0)$ – entropy coefficient of the systemic relationship between the lipid and hepatobiliary systems before treatment; $H(\mathbf{X}_1\cap\mathbf{Y}_1)$ – after treatment.

Based on a small number of observations (the sample was randomly reduced), preliminary data were obtained on different levels of statin safety in terms of effects on the hepatobiliary system – lovastatin had the worst indicators (this was the only non-original drug among the presented drugs). Standard statistical methods cannot detect this difference on such a large number of observations.

Table 1. Indicators of entropy coefficients of the systemic relationship between the lipid and hepatobiliary systems.

Medication	Entropy indicators			
	$H(\mathbf{X}_0 \cap \mathbf{Y}_0)$	$H(\mathbf{X}_1 \cap \mathbf{Y}_1)$	$\delta H(\mathbf{X} \cap \mathbf{Y})$	$\Delta H(\mathbf{X} \cap \mathbf{Y})$
Atorvastatin, *n*=14	0,403	0,135	33,4%	-0,269
Simvastatin, *n*=31	0,384	0,126	32,9%	-0,258
Fluvastatin, *n*=17	0,358	0,196	54,8%	-0,162
Lovastatin, *n*=11	0,443	0,644	145,6%	0,202

A promising method for studying population health is its modeling based on the risk analysis of a multidimensional stochastic system [17]. The method is based on the representation of a medical-biosocial system as a random vector with interdependent components controlled by its numerical characteristics. The use of this method allows us to quantify the health risk at the population level by determining the probability of adverse outcomes relative to the threshold and critical levels of each component (subsystem), as well as to assess the specific contribution of the latter. Risk analysis has a certain similarity to sea, since it also evaluates the dispersion and cross-correlation of components.

The probability of an unfavorable outcome P(D) (the probability of patients falling into the risk zone of diseases for the main biological risk factors relative to the threshold and critical levels of risk factors) is defined as

$$P(D) = P(\mathbf{X} \in D),\ D = \left\{ \mathbf{x} = (x_1, x_2, \ldots, x_m) : \sum_{j=1}^{m} \frac{(x_j - \theta'_j)^2}{b_j^2} \geq 1 \right\}.$$

P(D) was 0.262 in the group of healthy patients (25-34 years); in the group of patients with diseases: in the age group 18-24 years – 0.617; 25-34 years – 0.851; 35-44 years – 0.958. Thus, a single numerically expressed population estimate was obtained simultaneously for all major biological risk factors, taking into account their interdependence. At the same time, the method also allows you to determine the contribution to the assessment of each risk factor and, accordingly, choose a priority for prevention.

In the preventive work of a doctor, the main task is to identify patients with risk factors early, in the preclinical stage of diseases. "From a large number of people at low risk, a significantly higher number of cases of the disease may occur than from a small number of people at high risk" – the conclusion of the famous epidemiologist rose [18]. The risk analysis method makes it possible to detect patients with indicators near the upper limit of the norm, but within its limits by taking into account the mutual cross-correlation of these subsystems. With conventional methods of detecting one borderline level of risk factors without taking into account the multidimensional relationship (let's call this area a "dangerous" risk zone), such patients disappear from the area of preventive intervention (see table 2), for example, during medical examinations.

Table 2. Comparative analysis of detection methods for the main biological risk factors of patients in the "borderline" and "dangerous" risk zone.

Indicator	Age group, years					
	18-24	25-34	35-44	45-54	55-64	65+
The number of patients in the "dangerous" risk zone (the usual method of detection by 1 indicator), *n*	58	194	190	139	224	79
Number of patients in the "borderline" risk zone (based on risk analysis), *n*	87	249	208	149	237	81
Difference, *n*	29	55	18	10	13	2
Difference, %	50.0	28.4	9.5	7.2	5.8	2.5

5. Conclusion

Entropy-probabilistic modeling was used to analyze the system for monitoring factors in an organized population, in which preventive work was carried out in the framework of the who CINDI international program from 1994 to 2015. With multidirectional, unspoken changes in risk factors, it is difficult to give a unified assessment of the effectiveness of preventive measures and the state of population health. Using the entropy-probability model allows us to analyze the essence of changes occurring in the population at a different, deeper level and plan medical intervention. For example, in the entropy-probabilistic analysis of indicators that characterize the main risk factors, a young (18-44 years) male organized population in 2010 and 2015, the survey found: population health score P(D) in 2010 – 0.602, 2015 – 0.749; moderate negative changes (ΔP(D)=+0.147); - increased instability (ΔH=+4.9 %); - a slight increase in average levels of total cholesterol (by 8.2 %) and glucose (by 6.5%); - extremely low level of interaction between subsystems with a decrease in HR from - 0.1 to - 0.02; - the main contribution to the growth of P(D) was made by BMI and SAD with a decrease in the role of BMI and an increase in SAD.

We believe that the introduction of the entropy-probabilistic model in the educational process will help in the formation of the doctor's basic universal, professional competencies and systematic clinical thinking.

6. Funding

The work was supported by the RFBR (grant 20-51-00001 Bel_a).

References

[1] Mudrov M Ya 1949 *Word about the way to teach and learn practical medicine in the beds of patients* (Moscow USSR Academy of Medical Sciences) p 221
[2] Kewal K J 2014 *Textbook of Personalized Medicine* (New York: Springer Science+Business Media) p 732
[3] Glybochko P V, Svistunov A A and Shemetova G N 2008 General medical practice: stages of development and the role of the Saratov state medical University in its formation at the regional level *Saratov scientific and medical journal* **4(4)** 9-12
[4] Afanasyev V G 1981 *Society: system, knowledge and management* (Moscow: Politizdat) p 432
[5] Peregudov F I and Tarasenko F P 1989 *Introduction to system analysis* (Moscow: High school) p 367
[6] Nicolis G and Prigogine I 1977 *Self-organization in nonequilibrium systems: from dissipative structures to order through* fluctuations (Wiley) p 512
[7] Feinstein A R 1970 Pre-therapeutic classification of co-morbidity in chronic disease *Journal of Chronic Disease* **23(7)** 455-68
[8] Piepoli M F, Hoes A W, Agewall S et al 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation *Eur Heart J* **37(29)** 2315-81
[9] Prangishvili I V 2003 *Entropy and other systemic patterns: Issues of management of complex systems* (Moscow: Nauka) p 428
[10] Tyrsin A N, Kalev O F, Yashin D A and Lebedeva O V 2015 Population health assessment based on entropy modeling *Mathematical biology and bioinformatics* **10(1)** 206-19
[11] Tyrsin A. N. 2016 *Entropy modeling of multidimensional stochastic systems* (Voronezh: Science book publishing house) p 156
[12] Chen N et al 2020 Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study *The Lancet* **395** 10223

Journal of Physics: Conference Series **1691** (2020) 012162 doi:10.1088/1742-6596/1691/1/012162

[13] Nikiforov V V, Suranova T G, Chernobrovkina T Ya, Yankovskaya Ya D and Burova S V 2020 New coronavirus infection (COVID-19): clinical and epidemiological aspects *Archive of internal medicine* **10(2)** 87-93
[14] Yashin D A, Tyrsin A N, Kalev O F and Gevorgyan G. G. 2017 *System-entropy analysis of the effectiveness of lipid-normalizing drugs Modern problems of science and education* **6** http://www.science-education.ru/article/view?id=27296
[15] Jones P H, Kafonek S, Jaurora I et al 1998 Comparative dose efficacy study of atorvastatin versus simvastatin, pravastatin, lovastatin, and fluvastatin in patients with hypercholesterolemia (The CURVES Study) *Am. J. Cardiol.* **81** 582-87
[16] Tyrsin A N 2020 The relationship between entropy modeling and correlation analysis *Modern methods of the theory of boundary value problems: int. conf.* (Voronezh: "Science-Yunipress") 216-7
[17] Tyrsin A N, Kalev O F, Yashin D A and Surina A A 2018 risk model of a multidimensional stochastic system as a tool for studying the state of population health *System analysis and management in biomedical systems* **17(4)** 948-57
[18] Rose G 1985 Sick individuals and sick populations *Int. J. Epidemiol.* **14** 32-8